Conflict or Cooperation? Predicting Future Tendency of International Relations

Peng Chen
Kyoto University
Kyoto, Japan
chenpeng.acmer@yahoo.com

Adam Jatowt
Kyoto University
Kyoto, Japan
adam@dl.kuis.kyoto-u.ac.jp

Masatoshi Yoshikawa
Kyoto University
Kyoto, Japan
yoshikawa@i.kyoto-u.ac.jp

ABSTRACT

International relations analysis is crucial to many stakeholders including policy makers, executives in international companies or social scientists. Generally, recent events between two countries define the international relations between them. We explore the possibilities of predicting future tendency of international relations by analyzing historical events between countries. Using auto-coded event database GDELT (Global Data on Events, Location, and Tone), which records what happened between various countries in the past few decades, we extract various types of events between two countries of interest and aggregate them into categories: conflict and cooperation. Then, according to a sequence of recent events, we predict the number of conflict events and cooperation events in the next time unit. We use MILSTM (Multi-input LSTM) considering diverse kinds of relations between different country pairs. We assume that relations between a specific pair of countries could be affected by other related country pairs. Based on this hypothesis we first select country pairs related to the target pair, and extract their multiple historical event sequences as additional input to train the model. The test results show that MILSTM performs better than vanilla LSTM, which confirms our initial hypothesis.

KEYWORDS

Event Prediction, GDELT, News Collections, Time Series Analysis, Multi-input LSTM

ACM Reference Format:
Peng Chen, Adam Jatowt, and Masatoshi Yoshikawa. 2020. Conflict or Cooperation? Predicting Future Tendency of International Relations. In *The 35th ACM/SIGAPP Symposium on Applied Computing (SAC '20), March 30-April 3, 2020, Brno, Czech Republic.* ACM, New York, NY, USA, Article 4, 8 pages. https://doi.org/10.1145/3341105.3373929

1 INTRODUCTION

International relations define the way in which two or more nations interact with each other, especially in the context of political, economic, or cultural relationships. For example, more cooperation takes place when two countries have good relations, such as economic aid, military aid and judicial cooperation. For instance, the governments of the U.S. and Japan have agreed to further cooperation in space, including returning to the moon together, as SPACENEWS said on May 29, 2019[1]. This news gives testimony to the good relationship between Japan and the United States from the perspective of scientific and technological cooperation. In contrast, in another example, South Korea has filed a complaint with the WTO over Japanese trade restrictions, in the latest escalation of conflict between the two countries[2]. Obviously, a series of conflicts have had a negative impact on the relations between Japan and South Korea. Although international relations are complicated and multi-faceted, still one can model them with the help of recent events reported in the news.

Predicting has always been a central aspiration in the study of international relations, which means a lot to social scientists, journalists, policy makers and other stakeholders. It would be remarkable if a forecasting system is developed to give a reliable estimate of future tendency of international relations between countries of interest, giving a reference for journalists, policy makers and representatives of other related professions, helping them better understand the context of future international cooperation and conflict. In this paper, we predict the degrees of conflict and cooperation events which quantitatively characterize the relations between arbitrary two countries. In recent years, many attempts [2, 19, 20] have been made to predict future events. In the prediction tasks, one of the most challenging steps is data collection and building an applicable dataset for models. Fortunately, with the emergence of open-source online global event databases like ICEWS [8] and GDELT (Global Data on Events, Location, and Tone) [10], we can easily access structured data on news events from around the world. In particular, we use GDELT event table of Google Big Query for our experiments. It is worth mentioning that the GDELT dataset, one of the largest global event datasets, is freely available online, recording events from a variety of international news media with daily granularity. By extracting historical events between arbitrary two countries we can easily construct the sequence data we need.

The task after preparing the data is to build a model capable of processing and forecasting time series data. In the past few years, Recurrent Neural Networks (RNNs) have proved their powerful "memory ability" and have been used in a number of sequence-based learning problems [6]. Long Short-Term Memory (LSTMs), as a special kind of RNNs, is well-suited to learn long-term dependencies and has been widely used in many time series data processing tasks because of their outstanding performance. Oftentimes, LSTM serves as a building block for more advanced and complex models.

SAC '20, March 30-April 3, 2020, Brno, Czech Republic

ACM ISBN 978-1-4503-6866-7/20/03...$15.00
https://doi.org/10.1145/3341105.3373929

[1] https://spacenews.com/u-s-and-japan-to-cooperate-on-return-to-the-moon/
[2] https://www.ft.com/content/ea993216-d42d-11e9-8367-807ebd53ab77

In this paper, we develop the model based on Multi-input LSTM to predict future events between pairs of countries based on historical events, and compare its performance with that of vanilla LSTM. The purpose of developing MILSTM is also to prove a hypothesis that events occurring between other related country pairs could help improving the predictive performance of a specific pair of countries. In the experiments, we tried three different ways to choose related pairs for a given target country pair, and compared their performance.

The remainder of this paper is structured as follows. Section 2 introduces the related work. Section 3 describes the details of GDELT dataset used to extract event records between countries. Section 4 formalizes the prediction problem. Sections 5 and 6 demonstrate the models that inspired us and the structure of the model we proposed, respectively. In Section 7 we give the experimental settings and compare the performance of different models. Finally, Section 8 presents our conclusions and future work.

2 RELATED WORK

With the development of online news media, events occurring around the globe can be reported in near real time. GDELT is one such project that aggregates cooperation and conflict data from various newspaper sources [10]. Keertipati et al. [7] proposed a multi-level analysis of cooperation and conflict data in GDELT, giving an overview of how the GDELT project records events, and proved feasibility of using data extracted from GDELT to capture the global trends. Few works have studied how to make use of the GDELT dataset. Some traditional machine-learning methods have been introduced to do data mining in GDELT: Phua et al. [18] developed decision trees to predict the Singapore stock market's Straits Time Index using the GDELT dataset. Galla et al. [4] explored a series of methods, including Random Forest, Ada boost with random forest and LSTM, in the task of predicting social unrest using GDELT and GKG (Global Knowledge Graph). Qiao et al. [19] used the GDELT for predicting social unrest events across five major nations in Southeast Asia with Hidden Markov Model (HMM), and Yonamine et al. [24] utilized a statistical method, called Auto Regressive Fractionally Integrated Moving Average (ARFIMA), to predict violence levels in Afghanistan.

Smith et al. [20] explored the use of Neural Networks (RNNs and LSTMs) for predicting the number of conflict events in Afghanistan, compared with ARFIMA model. Wang et al. [22] proposed a more advanced model based on LSTM, called CALSTM (Context-Aware Attention LSTM), and achieved better prediction performance than LSTM. Besides, some previous works [5, 17, 23] have used other datasets for event prediction.

What all these works have in common is that they focus on one region or one country at a time. For instance, Yonamine et al.'s work focuses on violence levels in Afghanistan, Qiao et al.'s model ran on a data set of one country in southeast Asian at a time. Our work predicts cooperation and conflict between any pair of countries trying to estimate future tendency of international relations between different countries. Moreover, when predicting future events, we not only consider the target country pair, but also utilize information about other related country pairs during the same time period.

3 GDELT DATASET

GDELT (Global Data on Events, Location, and Tone) [10] is a real-time network diagram and database of global society, which monitors the world's news media from nearly every country in print, broadcast, and web formats, in over 100 languages. Despite certain problems and controversies [9, 14], GDELT remains largest resource of its type allowing to obtain broad coverage of world news.

GDELT event table is one of the cores of the GDELT project and stores more than 300 categories of events in CAMEO format (conflict and mediation event observations), extracted from the world's news media since 1979 by using natural language and data mining algorithms. Events are divided into four major categories. Each major category is made up of 5 root classes. Furthermore, there are several sub-classes following each root class. Thus, we have 20 event root classes and more than 300 event classes. In addition to the event type, each event record contains other attributes, such as the two actors involved in the event, date, number of mentions in all source documents, source URL and so on. Each event forms a row in the event table. Table 1 shows some examples with a subset of the event attributes.

Table 1: Examples of event records in GDELT.

SQLDATE	Actor1Code	Actor2Code	EventCode	QuadClass	GoldsteinScale	NumMentions
20150313	USA	JPN	43	1	2.8	4
20160919	USA	JPN	20	1	3	6
20170519	USA	JPN	40	1	1	4
20170519	USAGOV	JPN	61	2	6.4	8
20180517	USA	JPN	20	1	3	1
20181231	USA	JPN	193	4	-10	7
20181231	USA	JPN	80	2	5	2

SQLDATE is the date of event. Actor1Code and Actor2Code indicate two actors involved in the event, whose encoding format follows United Nations Country Codes[3], and event is the action performed by Actor1 on Actor2. In addition to ActorCode, there are other omitted attributes, such as ActorCountryCode, ActorName, etc. EventCode gives a fine-grained category of the event. QuadClass denotes major category the event falls under: (1) Verbal Cooperation, (2) Material Cooperation, (3) Verbal Conflict and (4) Material Conflict. More specifically, material cooperation or conflict can be one of the following types of events listed in Table 2.

Table 2: Examples of Material Cooperation or Conflict.

Material Cooperation	Material Conflict
Cooperate economically	Reduce or stop economic assistance
Provide economic aid	Halt negotiations
Provide military aid	Impose administrative sanctions
Share intelligence or information	Seize or damage property
Engage in judicial cooperation	Impose blockade
Ease political dissent	Use conventional military force
Ease military blockade	Attempt to assassinate

[3]https://unstats.un.org/unsd/tradekb/Knowledgebase/50347/Country-Code?Keywords=All

Each event is also assigned a numeric score ranging from -10 to +10 (negative for conflicts and positive for cooperation) which captures the theoretical potential impact that a given type of event has on the stability of a country. This is known as the Goldstein Scale. Finally, NumMentions is the total number of mentions in all source documents, which to some extent reflects the importance of the event.

As we focus on events occurring between two countries, the two actors of each event should be two different countries. We used ActorCountryCode to identify countries, NumMentions to filter out important events, and QuadClass to classify events. Finally, SQLDATE was used to get the time of the event.

In Figure 1, we counted the number of times each country acts as an event actor.

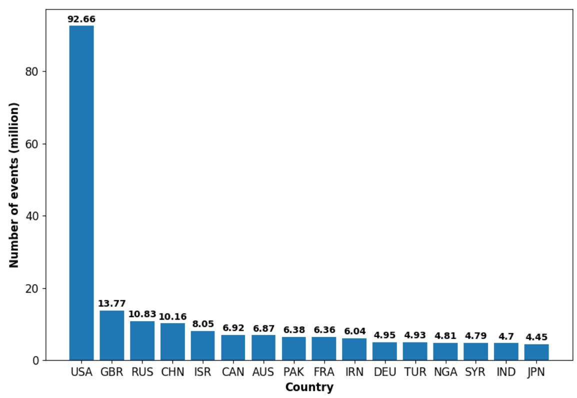

Figure 1: Country ranking by the number of events (16 countries).

Based on the figure above, 8 eventful countries were selected as the countries of interest as shown in Table 3. We then focus on the international relations between them.

Table 3: The set of analyzed countries.

Code	Country	Code	Country
USA	United States	GBR	United Kindom
RUS	Russia	CHN	China
CAN	Canada	AUS	Australia
FRA	France	JPN	Japan

4 PROBLEM STATEMENT

As mentioned before, our purpose is to predict future tendency of international relations between countries. Intuitively, the notion of international relations is rather vague and difficult to be quantified. We decided to predict the number of conflict and cooperation events in the future, which quantitatively characterizes the tendency of relations between two countries.

Given a country pair, we used one week as a time unit[4], aggregating the whole event sequence data by weeks. Specifically, the numbers of occurrences of events of four major categories (i.e., QuadClasses) at every week were counted, resulting in a vector $\vec{A_i} = \{c1, c2, c3, c4\}$, where c_k indicates the number of occurrences of events in category k in the i-th week. In other words, we counted the numbers of verbal cooperation, material cooperation, verbal conflict and material conflicts for every week. The whole sequence data for a pair of countries becomes $S = \{\vec{A_1}, \vec{A_2}, \vec{A_3}, \ldots, \vec{A_N}\} \in R^{N \times 4}$. Figure 2 and Figure 3 show some examples of the derived sequence data.

One issue that needs to be highlighted is that every event has its direction. If actor1 of the event is JPN, and actor2 is USA, the event belongs to the country pair (JPN, USA) (and not to (USA, JPN)). Therefore, each country pair represents a unidirectional relation.

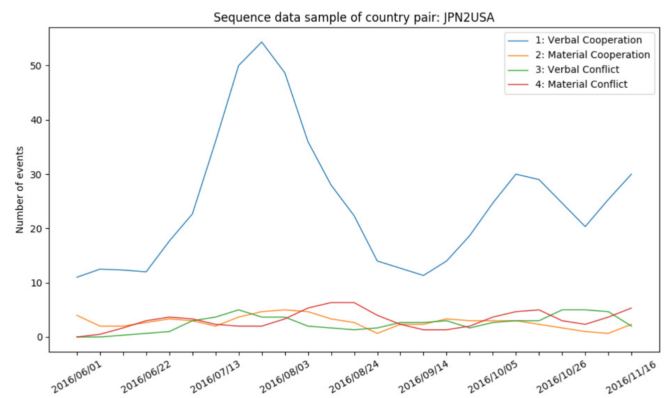

Figure 2: Sequence data of (JPN, USA) from 2016/06/01 to 2016/11/16.

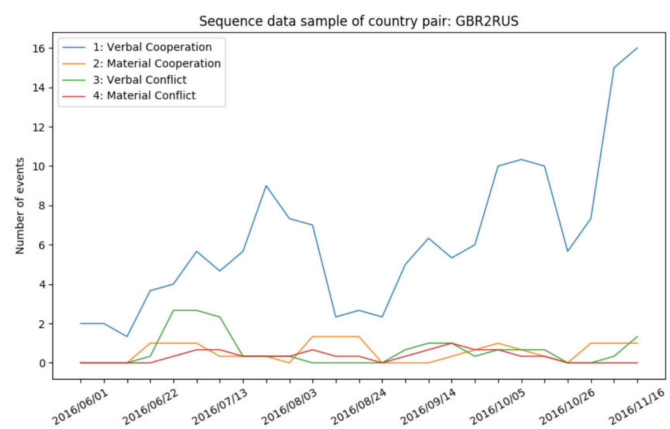

Figure 3: Sequence data of (GBR, RUS) from 2016/06/01 to 2016/11/16.

The prediction task can be defined as: Given the sequence data for consecutive M weeks $S_{(1,2,\ldots,M)} = \{\vec{A}_{t-M}, \vec{A}_{t-M+1}, \ldots, \vec{A}_{t-2}, \vec{A}_{t-1}\}$, predict the data for the (M+1)th week $\vec{A_t}$, i.e., the next week in the future. By this, we can investigate the tendency of international relations based on the numbers of conflicts (verbal and material),

[4]The choice of a week is to strike a balance between the size of a dataset and appropriateness of forecasting where the length of time unit should on average have sufficient number of events.

the number of cooperative events (also verbal and material) and their rates of change.

5 METHODOLOGY

5.1 Long Short-Term Memory Network

Recurrent Neural Networks (RNNs) are capable of processing sequential data with variable length by recursively applying a single transition function on hidden states. LSTMs are explicitly designed to avoid the long-term dependency problem that exists in standard RNNs [15].

There are five components in an LSTM unit: an input gate i_t, a forget gate f_t, an output gate o_t, a memory cell c_t, and a hidden state h_t. They are all vectors in $\mathbb{R}^k$, where k is the dimension of hidden state. Formally, given a time series $\{x_1, x_2, ..., x_T\}$ with $x_t \in \mathbb{R}^m$, each step in LSTM's recursive process can be defined as a collection of transition functions as follows:

$$
\begin{aligned}
i_t &= \sigma(W_i[h_{t-1}; x_t] + b_i) \\
f_t &= \sigma(W_f[h_{t-1}; x_t] + b_f) \\
o_t &= \sigma(W_o[h_{t-1}; x_t] + b_o) \\
u_t &= \tanh(W_u[h_{t-1}; x_t] + b_u) \\
c_t &= i_t \odot u_t + f_t \odot c_{t-1} \\
h_t &= o_t \odot \tanh(c_t)
\end{aligned}
$$

where x_t is the current input at a time step t, $h_t, h_{t-1} \in \mathbb{R}^k$ are the hidden states at time t and $t-1$, respectively. $W_i, W_f, W_o, W_u \in \mathbb{R}^{k\times(k+m)}$ are the weight matrices and $b_i, b_f, b_o, b_u \in \mathbb{R}^k$ are the bias vectors. σ refers to sigmoid function, $\odot$ denotes element-wise multiplication, and tanh represents hyperbolic tangent.

As it can be seen, when an input vector x_t comes in, the input gate decides which values in the input will be updated, then the forget gate controls how much the previous information from c_{t-1} is forgotten, and the output gate filters the memory state c_t, creating the next hidden state h_t.

For simplicity, we represent the transition of one LSTM layer as a single non-linear function f:

$$h_t = f(h_{t-1}, x_t)$$

Given a time series $S = \{x_1, x_2, ..., x_T\}$ with $x_t \in \mathbb{R}^m$, the output of LSTM is defined as:

$$S' = LSTM(S) = \{x_1', x_2', ..., x_T'\},\ t \in \{1, 2, ..., t\},\ x_t' \in \mathbb{R}^k$$

5.2 Attention Mechanism

The idea of attention mechanism is to assign attention weights on some layers of neural networks, and it has achieved a great success in many machine learning tasks. The intuition behind attention is that different parts in a given input should have different importance in generating the result sentence. Bahdanau et al. [1] employed attention in RNNs, assigning attention weights on different hidden states corresponding to different outputs. Context-Aware Attention LSTM [22] used attention mechanism on hidden states generated on different time steps. In the unit MILSTM that will be introduced later, the attention layer decides attention weights over all cell state inputs $l_t, l_{pt}, l_{nt}, l_{it}$, and aggregates them into L_t by weighted sum.

5.3 Multi-input LSTM

Li et al. [11] proposed attention-based Multi-input LSTM (MILSTM) for stock price prediction, taking into account the prices of other related stocks. This work inspired us to make use of historical data of related country pairs. Intuitively, international relations between two countries at a certain time depend also on relations between other (especially related) countries as well as the major events in the World at that time. The problem to be solved is however how to choose related country pairs given a target pair.

One limitation of the conventional LSTM architecture is that it can only handle a single sequence of data. Recently, several structural variants of LSTMs have been proposed. A tree-structured LSTM was introduced by Tai et al. [21], in which a LSTM unit may have multiple precedents. Liang et al. [12] first proposed graph LSTM for semantic object parsing in the field of image processing, which extends the traditional LSTMs from sequential data learning to general graph-structured data learning. Peng et al. [16] extended the use of graph LSTM to cross-sentence relation extraction. However, after preliminary experiments on GDELT using graph LSTM, we found that it is no better than vanilla LSTM in our prediction task.

The structure of a MILSTM unit is showed as follows:

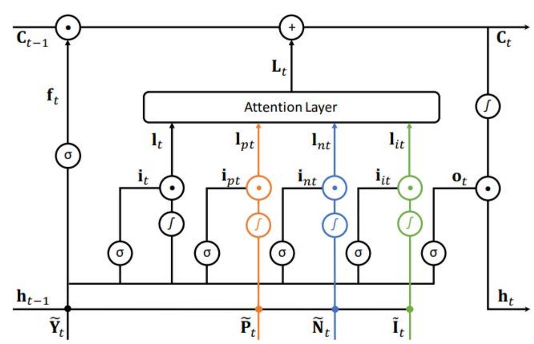

Figure 4: A MILSTM unit [11].

where Y is the target sequence, $\tilde{Y} = LSTM(Y)$ is the sequential embedding of Y, and $\tilde{Y}_t$ becomes the item of $\tilde{Y}$ at time step t. $\tilde{P}$ represents positively correlated sequence, while $\tilde{N}$ represents negatively correlated sequence. $\tilde{I}$ is index sequence, which is useful in stock prediction. $\tilde{Y}_t, \tilde{P}_t, \tilde{N}_t, \tilde{I}_t \in \mathbb{R}^m$.

For transition functions of MILSTM unit, the forget gate and output gate remain the same as vanilla LSTM:

$$
\begin{aligned}
f_t &= \sigma(W_f[h_{t-1}; \tilde{Y}_t] + b_f) \\
o_t &= \sigma(W_o[h_{t-1}; \tilde{Y}_t] + b_o)
\end{aligned}
$$

The cell states inputs transition becomes:

$$u_t = \tanh(W_u[h_{t-1}; \tilde{Y}_t] + b_u)$$
$$u_{pt} = \tanh(W_{pu}[h_{t-1}; \tilde{P}_t] + b_{pu})$$
$$u_{nt} = \tanh(W_{nu}[h_{t-1}; \tilde{N}_t] + b_{nu})$$
$$u_{it} = \tanh(W_{iu}[h_{t-1}; \tilde{I}_t] + b_{iu})$$

where $u_{pt}, u_{nt}, u_{it} \in \mathbb{R}^k$ (k is the dimension of hidden state), are the cell state inputs of target sequence, positively correlated sequence, negatively correlated sequence and index sequence, respectively.

$W_u, W_{pu}, W_{nu}, Wiu \in \mathbb{R}^{k\times(k+m)}$ are the weight matrices and $b_u, b_{pu}, b_{nu}, b_{iu}$ are their biases.

The input gate transition becomes:

$$i_t = \sigma(W_i[h_{t-1}; \tilde{Y}_t] + b_i)$$
$$i_{pt} = \sigma(W_{pi}[h_{t-1}; \tilde{P}_t] + b_{pi})$$
$$i_{nt} = \sigma(W_{ni}[h_{t-1}; \tilde{N}_t] + b_{ni})$$
$$i_{it} = \sigma(W_{ii}[h_{t-1}; \tilde{I}_t] + b_{ii})$$

The next step is like the first half of a transition function in vanilla LSTM $c_t = i_t \odot u_t + f_t \odot c_{t-1}$, using the input gates to filter the cell state inputs:

$$l_t = u_t \odot i_t$$
$$l_{pt} = u_{pt} \odot i_{pt}$$
$$l_{nt} = u_{nt} \odot i_{nt}$$
$$l_{it} = u_{it} \odot i_{it}$$

Then, the attention layer calculates the attention weights of each filtered cell state inputs:

$$r_t = \tanh(l_t^T W_a c_{t-1} + b_a)$$
$$r_{pt} = \tanh(l_{pt}^T W_a c_{t-1} + b_{pa})$$
$$r_{nt} = \tanh(l_{nt}^T W_a c_{t-1} + b_{na})$$
$$r_{it} = \tanh(l_{it}^T W_a c_{t-1} + b_{ia})$$
$$[a_t, a_{pt}, a_{nt}, a_{it}]^T = Softmax([r_t, r_{pt}, r_{nt}, r_{it}]^T$$
$$L_t = a_t l_t + a_{pt} l_{pt} + a_{nt} l_{nt} + a_{it} l_{it}$$

Finally, the cell state gets updated and generates hidden state:

$$c_t = c_{t-1} \odot f_t + L_t$$
$$h_t = \tanh(c_t) \odot o_t$$

Similarly, we can represent the transition functions of a MILSTM unit as the following function F and define the MILSTM function:

$$h_t = F(h_{t-1}, \tilde{Y}_t, \tilde{P}_t, \tilde{N}_t, \tilde{I}_t)$$
$$\tilde{Y}' = MILSTM(\tilde{Y}, \tilde{P}, \tilde{N}, \tilde{I})$$

This model presents us an architecture capable of handling multiple sequence learning problems. Besides, we are able to separate correlated sequences into two kinds: positive and negative, which makes the model flexible.

6 MODEL

In our proposed model, there are three main components: (1) Related pairs selection, (2) LSTM Encoder and (3) Multi-input LSTM.

6.1 Related Pairs Selection

Given a target country pair Y, we try to find top k positively related country pairs $\{X_1, X_2, ..., X_k\}$, which are likely sharing similar patterns of international relations. Y is composed of any two of 8 countries in Table 3. Note that the number of countries in the related country pairs is larger. $X_i \in \mathbb{X}$, where $\mathbb{X}$ is a collection of country pairs formed by 16 countries listed in Figure 1.

During the training phase, we can only access the training set. Suppose all the sequence data in the training set fall in the time period of $[T_{start}, T_{end}]$, and the whole training sequence data of each country pair can be represented as: $S = \{\vec{A}_1, \vec{A}_2, \vec{A}_3, ..., \vec{A}_T\} \in R^{T\times 4}$. Therefore, given a target country pair Y, and its data S_Y, the objective of selection is to find k country pairs with the highest relation by calculating $Rel(S_Y, S_{X_i})$, $X_i \in \mathbb{X}$ and $X_i \neq Y$. In this paper, we propose three methods to calculate the relation and we compare their performance in the experiments.

6.1.1 *Pearson Correlation Coefficient.*

Pearson Correlation Coefficient (PCC) is a measure of the linear correlation between two variables. We used PCC to measure the relation between sequence data of different country pairs. The PCC between two data sequences S_1 and S_2 is defined as:

$$Cor(S_1, S_2) = \frac{Cov(S_1, S_2)}{\sqrt{Var(S_1) * Var(S_2)}}$$

where Cov means covariance and Var means variance.

For the data sequence S of country pairs, there are 4 sub-sequences $S^{(j)}$, $j \in [1, 4]$, one for each major event category (i.e., QuadClass). We define the correlation between the two country pairs as follows:

$$Cor(S_Y, S_{X_i}) = \frac{1}{4}\sum_{j=1}^{4} \frac{Cov(S_Y^{(j)}, S_{X_i}^{(j)})}{\sqrt{Var(S_Y^{(j)}) * Var(S_{X_i}^{(j)})}}$$

The greater the correlation, the more likely the two country pairs should share an identical tendency. Table 4 shows few examples.

Table 4: Examples of the top 3 related pairs by PCC.

country pair	1st related	2nd related	3rd related
(JPN, USA)	(USA, JPN)	(USA, GBR)	(CAN, USA)
(CHN, USA)	(USA, CHN)	(GBR, USA)	(CAN, GBR)
(RUS, USA)	(USA, RUS)	(DEU, RUS)	(CHN, USA)

Given a target country pair, we find k related pairs with the largest k correlation values. If the target pair is (JPN, USA), we can find the top 3 related pairs from its correlation graph as in the figure below:

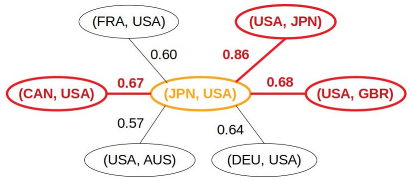

Figure 5: Top 3 related pairs of (JPN, USA) by PCC, marked in red.

6.1.2 *Geographical Distance.*

The intuition behind the next measure is that there is a chance that some countries share similar international relations' tendency if they are geographically close. Such countries may conduct similar foreign policy or may affect each other. We use the geographical distance between capital cities as a proxy of geographical distance between countries. Capitals are commonly the centers of national politics and places where key political decisions are typically made. The data source for the coordinates of capitals is taken from a public dataset on Kaggle, called "World capitals gps"[5].

To calculate distance between two country pairs, we only consider pairs that share one common country. For example, for a target pair (JPN, USA), another pair such as (USA, CHN) would be considered but not a pair such as (CHN, GBR). In the former case, we would calculate the geographical distance between the capitals of Japan and China, i.e.., Tokyo and Beijing, respectively. The following table shows some selection results.

Table 5: Example of top 3 related pairs by geo-distance.

country pair	1st related	2nd related	3rd related
(JPN, USA)	(USA, JPN)	(JPN, CAN)	(USA, KOR)
(CHN, USA)	(USA, CHN)	(CHN, CAN)	(USA, PRK)
(RUS, USA)	(USA, RUS)	(RUS, CAN)	(USA, UKR)

6.1.3 *Semantic Similarity.*

We hypothesize that country pairs which have high semantic similarity to the target pair have high chance to be involved in similar or related events. We then calculate the cosine similarity of the semantic vectors (word2vec [13]) of country names. By utilizing pretrained word embedding (word2vec on Google News[6]), we obtained semantic vector of each country name, and defined the semantic vector of a country pair as the difference between the corresponding vectors of the two countries. For example, the semantic vector of (JPN, USA) is defined as $\overrightarrow{JPN, USA} = \overrightarrow{USA} - \overrightarrow{JPN}$, whose dimension is 300. The relation is computed as follows:

[5] https://www.kaggle.com/nikitagrec/world-capitals-gps
[6] https://code.google.com/archive/p/word2vec/

$$\begin{aligned} Cor(Y, X_i) &= cos(\overrightarrow{N_Y}, \overrightarrow{N_{X_i}}) \\ &= \frac{\overrightarrow{N_Y} \cdot \overrightarrow{N_{X_i}}}{||\overrightarrow{N_Y}|| \, ||\overrightarrow{N_{X_i}}||} \end{aligned}$$

where $\overrightarrow{N_Y}$ and $\overrightarrow{N_{X_i}}$ denote the semantic vectors of the country names of Y and X_i, respectively. In Table 6, we show some selection examples.

Table 6: Example of top 3 related pairs by semantic similarity.

country pair	1st related	2nd related	3rd related
(JPN, USA)	(CHN, USA)	(DEU, USA)	(AUS, USA)
(CHN, USA)	(IND, USA)	(JPN, USA)	(RUS, USA)
(RUS, USA)	(TUR, USA)	(CHN, USA)	(IRN, USA)

6.2 LSTM Encoder

For the target country pair and its related country pairs, we first fed their sequence data $S_Y, \{S_{X_1}, S_{X_2}, ..., S_{X_k}\}$ into LSTM, encoding them into sequential embeddings which are hidden states generated in all time steps.

$$\begin{aligned} S'_Y &= LSTM(S_Y) \\ S'_{X_i} &= LSTM(S_{X_i}), i \in [1, k] \end{aligned}$$

Note that the LSTMs in the above equations share parameters.

In order to reduce the size of input variables for Multi-input LSTM, all sequential embeddings of the related country pairs are aggregated into one sequence by averaging:

$$S'_X = \frac{1}{k} \sum_{i=1}^{k} S'_{X_i}$$

Therefore, there are two sequential data, S'_Y and its related auxiliary, positively correlated data S'_X, as the input to the next component.

6.3 Multi-input LSTM

In our model, we do not use negatively correlated sequence and index sequence as is in the case of the original MILSTM. Thus, we developed an adapted version of original MILSTM. At each time step t, we take the output of LSTM encoder S'_{Y_t} and S'_{X_t} as the input to the model. The illustration of a model unit is shown below:

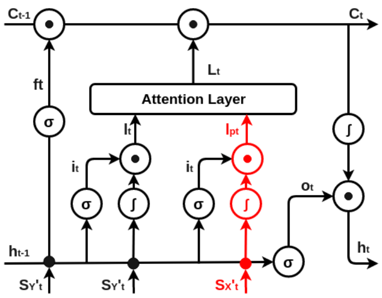

Figure 6: The structure of a unit in the model.

The transition process at each time step is similar to the one in the original MILSTM, but we have only one positively correlated input S'_X.

The workflow of the whole model can be illustrated as follows:

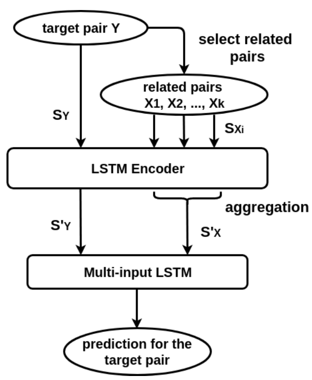

Figure 7: The workflow of our model.

7 EXPERIMENT

7.1 Experimental Settings

We first filtered out less important events for which the value of NumMentions is smaller than 3, and applied moving average smoothing on the remaining raw data with window size equal to 3. As a result, each vector $\vec{A_i} \in R^4$ in $S = \{\vec{A_1}, \vec{A_2}, \vec{A_3}, ..., \vec{A_N}\}$ becomes the average of previous 3 vectors including itself: $\vec{A'_i} = \frac{1}{3}\sum_{j=0}^{2} \vec{A_{i-j}}$, where the summation here means element-wise summation.

In the experiments, we first predict the number of material conflict events in the next time slot, as this kind of conflicts have a fatal and long-term effect on international relations between two countries. Furthermore, we also tested the performance of predicting material cooperation events.

As mentioned in Section 4, M represents the length of historical sequence data, and is set to 15. Therefore, in the experiment, we predict the number of material conflict events for the 16-th week $\vec{A_t}$ given the sequential data of the previous 15 weeks $S_{(1,2,...,15)} = \{\vec{A_{t-15}}, \vec{A_{t-14}}, \vec{A_{t-13}}, ..., \vec{A_{t-1}}\}$.

For the partition of dataset, the training set contains sequential data from 2005/01/01 to 2016/06/01 (595 weeks), while the test set ranges from 2016/06/01 to 2018/12/01 (130 weeks). The ratio of training to testing data is close to 8:2. Other parameters in our system include time unit = week, length of input sequence = 15, hidden size of LSTM encoder = 256, hidden size of MILSTM = 256, learning rate = 0.0002, number of training epochs = 40. Besides, we select top 3 positively related country pairs when using our model.

For the calculation of loss, we use mean square error (MSE) over all test samples as the loss metric.

7.2 Evaluation

The following figure illustrates an example of prediction of the number of material conflict on the whole test dataset when the target pair is (USA, CHN):

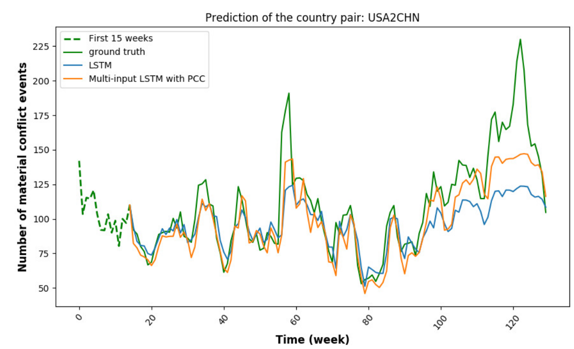

Figure 8: Prediction performed by LSTM (blue line) and by MILSTM with PCC (orange line). The green line denotes ground truth, and the dotted line represents the first 15 weeks. The prediction starts from the 16th week.

We evaluated Multi-input LSTMs using different selection methods for related country pairs based on 56 country pairs formed by any two of 8 countries of interest. Table 7 below shows MSE loss of prediction for selected country pairs as well as the average performance (the last row).

Table 7: MSE loss of prediction of material conflict events.

	LSTM	MILSTM with PCC	MILSTM with geo-distance	MILSTM with word2vec
(USA, GBR)	422.70	**310.54**	507.31	340.36
(USA, RUS)	8597.08	**5576.87**	6687.25	6415.67
(USA, CHN)	810.63	484.94	**481.73**	673.34
(USA, CAN)	339.25	359.96	**302.25**	336.09
(USA, AUS)	685.59	604.46	550.17	**512.03**
(USA, FRA)	188.89	195.96	**172.75**	212.80
(USA, JPN)	172.07	**157.20**	172.75	178.31
(GBR, USA)	683.35	549.93	500.72	**475.35**
(GBR, RUS)	8795.58	8414.33	**7729.35**	7782.66
(GBR, AUS)	57.23	56.58	**50.24**	59.88
(GBR, CAN)	**47.09**	51.83	48.55	59.88
(GBR, AUS)	57.23	56.58	**50.24**	52.02
(GBR, FRA)	102.82	90.38	**85.57**	183.67
(RUS, USA)	2643.09	**1690.20**	2257.40	1762.92
(RUS, CHN)	95.61	90.42	**88.08**	93.35
(JPN, USA)	910.20	**720.37**	784.59	747.14
....	....	....	....	....
Average improvement over LSTM		10.42%	11.69%	7.50%

The table below compares the performance of predicting different types of events, including material cooperation and material conflict.

Table 8: Average improvement of MILSTM over LSTM for predicting different types of events

event type	MILSTM with PCC	MILSTM with geo-distance	MILSTM with word2vec
material cooperation	4.36%	12.95%	6.52%
material conflict	10.42%	11.69%	7.50%

The average performance of LSTM is worse than that of Multi-input LSTM. Especially, in the case where the value of loss is large, LSTM usually does not perform well. Large loss is generally caused by a large number of events that occurred between two countries, where data tends to exhibit stronger regularity and learnability, so that our model is more likely to perform well, while irregular fluctuations are likely to occur when the amount of data is small.

In general we observe that adding contextual data in the form of recent events of related country pairs helps to boost the accuracy of results. In particular, selecting country pairs that are geographically close seems to result in the largest improvement.

Still, there is room for improvement for our model, such as changing the way of aggregation of related sequential inputs $\{S'_{X_1}, S_{X'_2}, ..., S_{X'_k}\}$, and choosing a more appropriate selection method, etc.

8 CONCLUSIONS & FUTURE WORK

In this paper we have explored the usage of GDELT dataset for forecasting cooperation and conflict events in international relations. We developed our model based on multi-input LSTM and on the selection of different related country pairs. Through experiments, we found that adding relevant information from related country pairs does indeed help the model achieve better performance.

Note that in the experiments, we tested some country pairs for which there is largest number of event records. The next step is to run the models on the entire set of country pairs. From the evaluation results, we found that introducing additional inputs not only results in more computation cost but sometimes may be harmful to the prediction of the target sequence. Hence, how to avoid related pairs becoming noise forms a part of our future work.

In the future, we will investigate more forecasting models. Nowadays, there are many solutions in the field of multivariate time series processing. Stock forecasting is one example, for which various advanced models have been proposed. Feng et al. [3] devised Temporal Relational Ranking system for stock prediction which could serve as a inspiration for subsequent improvement of our models. Specifically, they also considered the relations between different stocks and used Graph Convolutional Networks (GCN) to perceive information from related stocks. The first half of their model also used LSTM to encode the raw sequential inputs and to generate sequential embeddings, which provide node features to the GCN. Our future work includes exploring other kinds of models that are structurally suitable for analyzing data from GDELT.

REFERENCES

[1] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2014. Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473* (2014).

[2] Patrick T Brandt, John R Freeman, and Philip A Schrodt. 2011. Real time, time series forecasting of inter-and intra-state political conflict. *Conflict Management and Peace Science* 28, 1 (2011), 41–64.

[3] Fuli Feng, Xiangnan He, Xiang Wang, Cheng Luo, Yiqun Liu, and Tat-Seng Chua. 2019. Temporal relational ranking for stock prediction. *ACM Transactions on Information Systems (TOIS)* 37, 2 (2019), 27.

[4] Divyanshi Galla and James Burke. 2018. Predicting Social Unrest Using GDELT. In *International Conference on Machine Learning and Data Mining in Pattern Recognition*. Springer, 103–116.

[5] Jesse Hammond and Nils B Weidmann. 2014. Using machine-coded event data for the micro-level study of political violence. *Research & Politics* 1, 2 (2014), 2053168014539924.

[6] Andrej Karpathy. 2015. The unreasonable effectiveness of recurrent neural networks. *Andrej Karpathy blog* 21 (2015).

[7] Swetha Keertipati, Bastin Tony Roy Savarimuthu, Maryam Purvis, and Martin Purvis. 2014. Multi-level analysis of peace and conflict data in GDELT. In *Proceedings of the MLSDA 2014 2nd Workshop on Machine Learning for Sensory Data Analysis*. ACM, 33.

[8] Brian Kettler and Mark Hoffman. 2012. Lessons learned in instability modeling, forecasting, and mitigation from the DARPA integrated crisis early warning system (ICEWS) program. In *2nd International Conference on Cross-Cultural Decision Making: Focus*.

[9] Haewoon Kwak and Jisun An. 2016. Two tales of the world: Comparison of widely used world news datasets GDELT and EventRegistry. In *Tenth International AAAI Conference on Web and Social Media*.

[10] Kalev Leetaru and Philip A Schrodt. 2013. Gdelt: Global data on events, location, and tone, 1979–2012. In *ISA annual convention*, Vol. 2. Citeseer, 1–49.

[11] Hao Li, Yanyan Shen, and Yanmin Zhu. 2018. Stock Price Prediction Using Attention-based Multi-Input LSTM. In *Asian Conference on Machine Learning*. 454–469.

[12] Xiaodan Liang, Xiaohui Shen, Jiashi Feng, Liang Lin, and Shuicheng Yan. 2016. Semantic object parsing with graph lstm. In *European Conference on Computer Vision*. Springer, 125–143.

[13] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781* (2013).

[14] Kenneth McCubbins Nicholas Weller. [n. d.]. Raining on the Parade: Some Cautions Regarding the Global Database of Events, Language and Tone Dataset. https://politicalviolenceataglance.org/2014/02/20/raining-on-the-parade-some-cautions-regarding-the-global-database-of-events-language-and-tone-dataset/.

[15] Christopher Olah. [n. d.]. Understanding LSTM Networks. http://colah.github.io/posts/2015-08-Understanding-LSTMs/.

[16] Nanyun Peng, Hoifung Poon, Chris Quirk, Kristina Toutanova, and Wen-tau Yih. 2017. Cross-sentence n-ary relation extraction with graph lstms. *Transactions of the Association for Computational Linguistics* 5 (2017), 101–115.

[17] Chris Perry. 2013. Machine learning and conflict prediction: a use case. *Stability: International Journal of Security and Development* 2, 3 (2013), 56.

[18] C Phua, Y Feng, J Ji, and T Soh. [n. d.]. Visual and predictive analytics on Singapore news: experiments on GDELT, Wikipedia, andâĹğ STI (2014). *URL http://arxiv. org/abs/1404.1996 http://www. sas. com/singapore* ([n. d.]).

[19] Fengcai Qiao, Pei Li, Xin Zhang, Zhaoyun Ding, Jiajun Cheng, and Hui Wang. 2017. Predicting social unrest events with hidden Markov models using GDELT. *Discrete Dynamics in Nature and Society* 2017 (2017).

[20] Emmanuel M Smith, Jim Smith, Phil Legg, and Simon Francis. 2017. Predicting the occurrence of world news events using recurrent neural networks and auto-regressive moving average models. In *UK Workshop on Computational Intelligence*. Springer, 191–202.

[21] Kai Sheng Tai, Richard Socher, and Christopher D Manning. 2015. Improved semantic representations from tree-structured long short-term memory networks. *arXiv preprint arXiv:1503.00075* (2015).

[22] Xiuling Wang, Hao Chen, Zhoujun Li, and Zhonghua Zhao. 2018. Unrest News Amount Prediction with Context-Aware Attention LSTM. In *Pacific Rim International Conference on Artificial Intelligence*. Springer, 369–377.

[23] Nils B Weidmann and Michael D Ward. 2010. Predicting conflict in space and time. *Journal of Conflict Resolution* 54, 6 (2010), 883–901.

[24] James E Yonamine. 2013. Predicting future levels of violence in afghanistan districts using gdelt. *Unpublished manuscript* (2013).